

Entry

Large Language Models and Logical Reasoning

Robert Friedman [†] 

Department of Biological Sciences, University of South Carolina, Columbia, SC 29208, USA;
bob.network.science@gmail.com

[†] Retired.

Definition: In deep learning, large language models are typically trained on data from a corpus as representative of current knowledge. However, natural language is not an ideal form for the reliable communication of concepts. Instead, formal logical statements are preferable since they are subject to verifiability, reliability, and applicability. Another reason for this preference is that natural language is not designed for an efficient and reliable flow of information and knowledge, but is instead designed as an evolutionary adaptation as formed from a prior set of natural constraints. As a formally structured language, logical statements are also more interpretable. They may be informally constructed in the form of a natural language statement, but a formalized logical statement is expected to follow a stricter set of rules, such as with the use of symbols for representing the logic-based operators that connect multiple simple statements and form verifiable propositions.

Keywords: large language models; deep learning; symbolic logic; propositional logic; logical reasoning



Citation: Friedman, R. Large Language Models and Logical Reasoning. *Encyclopedia* **2023**, *3*, 687–697. <https://doi.org/10.3390/encyclopedia3020049>

Academic Editors: Kamran Munir,
José Raúl Romero, Khalid Hafeez,
Gemine Vivone and Raffaele Barretta

Received: 12 May 2023
Accepted: 29 May 2023
Published: 30 May 2023



Copyright: © 2023 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Background

The large language models of deep learning depend on natural language samples from a large corpus of knowledge [1]. However, natural language is a communication form that is derived from biological evolution [2–4] and is a descendant of the other related forms of animal communication, a recent adaptation with roles in sociality, social organization, and general behavior in a natural context [5]. It is also not considered a rigorous method for communication of knowledge that includes the attributes of reliability and permanence [6,7]. Likewise, the common definitions of rhetoric include the act of effective writing—but as an art form, not as a scientific practice [8,9]. This observation suggests that natural language is limited in its capacity for the construction of reproducible knowledge, a problem studied in philosophy, such as in the literature of epistemology and logic [6].

This problem is also studied in the context of deep learning models in computer science [10–13]. For instance, Traylor and others [11] recorded visual data from experiments for whether the logical operations of negation [14] and disjunction [15] are learned by neural network models. They showed that these models acquire some generalized knowledge of these tasks, but are not yet reliable for generating the elements of logical reasoning. In another study, Evans and others [12] published a data set as a benchmark to test for logical entailment, a subtype of propositional logic that relies on symbolic-based reasoning [16,17]. These formalized methods of logic may be contrasted against the more informal methods of reasoning. However, the formality of propositional logic offers interpretability on the processes of logical reasoning [18–24].

A strict method of application of propositional logic is to show the symbols of logic as Boolean operators on statements of natural language. An example is negation of a statement [14]. Given the statement “human perception is innate”, in a classical context, the negation of this statement is formed by NOT as a Boolean operation on this statement. Therefore, the negation of the statement is “human perception is not innate”. For the case of disjunction in propositional logic, at least in a classical context, two simple statements are joined by the Boolean operator OR [15]. Therefore, an example is “human perception

is either innate or not innate". However, other natural languages may represent a formal disjunction using a different method, so it is necessary to contrast the pure forms of ideal logic with its use in common practice [15,24,25]. Logic is a form of reliable expression of concepts for compensation of vagueness and error in interpretation of statements across the natural languages. Therefore, training a machine system on samples of natural language is expected to reproduce these same errors. However, logic is taught, so it is possible to acquire knowledge of logical reasoning.

The alternative to these formalized approaches is to rely on metaphysical and non-mechanistic definitions of general and logical reasoning. A similar problem exists in the common definitions of "intelligence". These amorphously formed terms stem from concepts that precede the best practices in science and knowledge of the mechanics of cognition and perception. In this case, as for other natural processes, a mechanistic approach to the explanation of phenomena favors experimentation, reproducibility, and scientific rigor. A second alternative is to rely on definitions based on a naive perspective, and, therefore, potential exclusion of the mechanistic basis of natural science and any of the rigor from the methods of philosophy. Since the various schools and teachings of philosophy are not expected to influence common sense as a population phenomenon, it is likely that common sense is often formed by methods without a component of verifiability [26].

Deep learning and its models are a testbed for verification of information-based phenomena [27,28]. For example, studies have shown that these models are capable of generating machine-readable code from queries in a text-based prompt, such as in the form of natural language requests; therefore, this capability shows that logic is computable by the advanced large language models, as in GPT-4 [28,29]. Bubeck and others [29] further showed that logical reasoning in the GPT-4 model is observed in its solutions for mathematical and general reasoning problems. The higher-order capabilities of these models are often referred to as emergent properties that arise from scaling the model with the inclusion of "big data" sets [30].

It may be that large language models are incorporating samples of propositional logic during the training step, where the data source is a natural language corpus. However, these capabilities may be contrasted against training a pretrained model on a large number of samples with formal statements of propositional logic [25]. The presumption is that tokenization of these samples would better capture higher-order logic than "common speech". Another method is to connect external tools to the deep learning model using a translation layer that communicates natural language into a machine-readable format [31]. These exchanges between the model and external tools are information that is not incorporated into the neural network itself, but these exchanges may be recorded and added to the model during a later training phase.

2. Construction of Logical Statements

2.1. Deep Learning Models

The procedure to associate natural language samples with formal logical statements is essentially a translation from one language to another. The vanilla transformer architecture is adapted for this kind of translation [28], including in its use of an encoder for learning knowledge of the correspondence between a language sample and its translation in the form of a logical statement [18,32,33], although it is an open question on the utility of the formalism of symbolic logic, since propositional logic can be composed without symbols. The transformer also has a decoder part which is coupled with the encoder, so the encoded knowledge is transferred to the decoder which is capable of generating natural language. Lastly, this procedure is expected to lower the dimensionality inherent in natural language data, and likewise, there is an expectation for higher performance in identifying the generalizable features [19,24].

Large language models that include samples of logic may increase their reliability, parameterization of the forms of logic, and interrelatedness of higher-order concepts. Just as these language models are based on a network of artificial neurons, concepts and their meanings are also an abstract form of a network. Both examples are essentially communication systems since they move information from one node to another by a connection.

Propositional logic is defined as a “... branch of logic that studies ways of joining and/or modifying entire propositions, statements or sentences to form more complicated propositions, statements or sentences, as well as the logical relationships and properties that are derived from these methods of combining or altering statements” [18]. Therefore, this logic is applicable to the construction of complex ideas and descriptions of them. There is also an assumption that the elemental (simplest) sentences that combine to form these propositions are logically valid [18]. However, validation of the simple sentences may be approached as a separate problem from identification and deconstruction of the varieties of complex statements.

Natural language generally relies on an imprecise use of words and their meanings for expression of strict logic, such as when using a conjunction in combining statements with the word “and” or “but”. It is possible to form the same conjunction with either of these words. However, symbolic logic offers precision in the expression of logic in complex statements. The large language models may capture and represent logic by an extensive sampling of a natural language corpus, but these models are expected to show improvement where the training step includes a large number of samples along with corresponding translations in a logic-based format.

It is also possible to train a deep learning model specific for translating natural language samples to a logical format. In this case, the output may then be used as input for any pretrained large language model. This is akin to a modularized design in information processing.

2.2. Models of Tokenization

Another problem is in the data samples of natural language. Typically, tokens are assigned to subwords for converting a natural language corpus to a machine-readable format in deep learning. The layers of the neural network are capable of capturing the higher-order representations in data [34], such as the semantic meaning of words in a paragraph or the interrelatedness of ideas in a corpus. However, this training process is not considered robust, since the large language models may generate output that does not mimic common speech. In the context of human interpretation, this is often described by an impression of output that is wrong or incorrect, although there are at least two causes for this interpretation, one is that the output is not interpretable, such as from insufficient data samples, or a limitation of the model itself, while the other is that the human interpreter is stating a mere opinion with a high degree of certainty [6].

In the case of scientific data and application, deep learning is mainly free of the subjectivity of human speech in the language models, but this application has insights into the kinds of approaches for tokenizing data. The AlphaFold project [35] developed a model for predicting the three-dimensional structure of cellular proteins. They used a sequence model—a transformer architecture for the major computational step [28], and this step includes data beyond the identification of amino acids in the protein sequence, such as in use of the biochemical and geometric features inherent in the sequence data. This is one approach. It also depends on a complex pipeline of processing biological data. Instead, ESMFold [36] confined the model to the amino acid composition at the sequence level without the appending of information on higher-order features such as geometrical arrangements at the atomic level. Their model showed a high capability in generating the three-dimensional structure of cellular proteins from sequence data, where the model was scaled in size along with the inclusion of very large data sets in a format that incorporates information on evolutionary relatedness [36].

This is a case where data and the data types are adaptable to a tokenization process. However, this process is not expected to fully capture the molecular dynamics of cellular proteins, since these molecules undergo conformational changes over time, particularly in their dynamics and interactions with other molecules. For this case, data on the temporal dynamics of molecules are required in some form.

Natural language data has the same limitations. A generative model is dependent on sufficient data sampling for constructing higher-order representations. The tokens are directly dependent on a sampling process and that are tractable for study in overcoming the limitations of a model. If a large language model is not reproducing logic from an existing corpus, it is reasonable to first scale the model upwards, but the hypothesis should then be posed on whether the appropriate tokens are present for training the model, including the breadth in data sampling, and specifically how the data is tokenized for putative formation of the higher-order features.

2.3. Prompt-Based Methods in Deep Learning

In the case of a large language model, the use of propositional logic may increase the interpretability of the model's output with the goal of finding the causes of unreliability in the model [37]. A related approach is a procedure for prompting an advanced large language model for directing the model to align the output with that of an expectation [37–41]. In particular, Creswell and others [37] formalized a procedure to train the neural network on a type of step-by-step reasoning process. This allows for output that is more interpretable, since each step is disentangled for analysis. Their study [37] also showed that the model is fairly robust in the generation of simple statements, at least with use of additional prompting, so the sharp limitations of the large language models are expected to mainly occur in complex statements of logic.

These kinds of models—based on prompting or pretraining at the prompt—are capable of logical reasoning by criteria; therefore, one approach is to automate away the multiple prompting procedure, while another method is to assign a certainty score attached to the model's generative output. Probability scores can be derived internal to the model and its computations if the expectations are explicit, such as with the use of symbolic logic in the formal expression of concepts [23,24], while another is to have a judge that is external to the model, such as by verification through the use of external tools. It is also possible to combine these approaches, where the exchanges between the model and external tools are subsequently trained into the model itself. In any case, it is important for the model to identify statements as a probability of truthfulness by a criterion, and where uncertain, generate output with the reasons for uncertainty.

2.4. Validation of Models

It is important to differentiate terminology based on metaphysics from that which is scientifically verifiable in the context of a deep learning system. Terms should depend on mechanistic science that is free of a metaphysical basis. In addition, the models themselves should be studied in a scientific context for the advancement of knowledge [10]. Many of the recent advanced large language models are used for their variability in the generation of output and capability for imitating human appealing speech. However, for the purpose of experimentation, the generative output should not depend on a high degree of variety and aesthetics, but instead on validity and reproducibility [42]. This lowers the number of external factors in a controlled experiment. An example of this is to switch to use of a “greedy search” by the sequence decoder in a large language model, so an excess of output variability is not a confounding factor.

In the case of general reasoning in the large language models, it is a good practice to describe hypotheses that are scientific and based on a mechanistic view, so the elements of general reasoning are clearly defined. Furthermore, if model scaling does not capture these elements from common samples, then propositional logic [18] and the tenets of science are paths for enhancing a model's reliability. Proper hypothesis testing, preferably with falsifiable statements, is expected to reduce dependence on nonscientific description and unverifiability of findings. However, the commercialization of these models is a separate problem, with reliance on methods outside the tenets of scientific investigation.

3. Problems in Logic and Language

3.1. Internal Representations of Logic

Computational semantics [23], the bridge between the study of language and logic, is a theoretical basis for predicting the capabilities and limitations inherent in a large language model. Liang and Potts [23] reviewed this area of knowledge with a focus on two lineages of study; one strives to map language upon the distinct forms of logic, while the other pursues a distributed approach, as in the vector representations in deep learning, so as to identify the mechanisms for natural language (re)cognition. They also introduce the concept of generalization as a major tenet for the study of meaning and the structural complexity in natural language.

Generalization is an insightful point, because without this property, it is expected that knowledge would not be attainable, so instead, each problem in semantics would be unique and distinct in origin. The concept of “logical forms” offers a theoretical basis [23], both as a formalized description of semantics, and as an overarching principle, for hypothesizing on the concept of meaning that is expressed in natural language [24]. The concept of meaning is deconstructed in a mechanistic context by Liang and Potts [23], but a naive viewpoint on the meaning in natural language may instead converge on solely a metaphysical view, such as that meaning originates in mental processes ungrounded by the rules of the physical world. Given a scientific viewpoint, the logical forms are expected to follow the rules of information theory, with a dependence on information flow as a physical phenomenon, and that these logical forms are expectedly resistant to information compression [43]. This is unlike sensory objects as formed by percepts that originate in the physical world. These latter forms are expectedly compressible, as observed in the deep learning models and their capability for generalization in sampling visual data [44].

Mathematics is a different kind of language in that it relies on symbolic forms, precision in meaning, and is lower in dimensionality than natural language. In this case, it is simpler to show that meaning is derived from a set of learned rule sets, as observed in the symbolic representations of mathematical concepts [45]. If a math problem is formulated by symbols and their expressions, then these calculations are programmable. If they are programmable in a computer language that translates to machine code, then a deep learning system can train on and learn these rules [5,46,47]. This contrasts with natural language where the rules are imprecise and not strictly bounded or defined. However, the scaling of large language models, leading to very large parameter sizes, has led to capturing the structural elements of natural language, such as the grammar rules, and the capability of generating common samples of speech.

3.2. Potential Limitations of Logical Systems

There is an observation of brittleness in the large language models—that they are not fully capturing the meaning of concepts [48]. This is not a result of a deficit in emulating a mental process (non-physical phenomena), but instead a lack of data and the programming of the model.

There are at least two hypotheses on large-scale language models. One is that scaling, including the size of the neural network, will lead to capturing the higher-level forms that are inherent in natural language. The other hypothesis is that the computational complexity will rapidly tend toward infinity, so progression in understanding, such as by complex logic, would be limited by theoretical constraints on use of data and computational time. Since the human brain is capable of finding meaning in natural language, support for the latter hypothesis would suggest that any limits in deep learning are instead one of optimization of methodology.

Using empirical findings and information theory as a guide [43], it is expected that internal cognition is processing information at a high dimensional level (uncompressed) and that its coding is organized in a heterogeneous and disordered fashion in the system [49]. One hypothesis is that natural language sources are insufficient for learning and making robust predictions about the world, but instead visual information—at least visual representations—are required for robust predictions on the properties and interactions of objects in the world, including those as perceived by the senses and those constructed internally by pure cognition. An example of the latter is a mathematical object, such as an arithmetic operator.

A second hypothesis, not exclusive to the first one, is that meaning is recognized by processes that rely on spatial scale across natural language samples. It is presumed that the higher-scale representations encoded by natural language are captured in the many layers of the neural network in deep learning, but it is unknown whether tokenization of natural language samples by the use of subwords is completely efficient. It may be that these neural networks do not have the data nor network scale to robustly compute these higher-order representations from current tokenization practices for natural language. In this case, a model that can process a hierarchy of tokens at different spans across samples may be considered. An analogy is in the construction of the monumental pyramids in ancient Egypt. If they were constructed of small blocks, then it would be a more difficult assembly process than the actual construction, which is based on assembly by very large blocks [50]. It follows that it would be difficult for a large language model to capture meaning in natural language samples from tokenization at the character level. It does capture meaning at least at a subword level, however, since these tokens have correspondence to the property of “meaning”, and the models are capable of capturing the higher-level representations from this procedure (Figure 1).

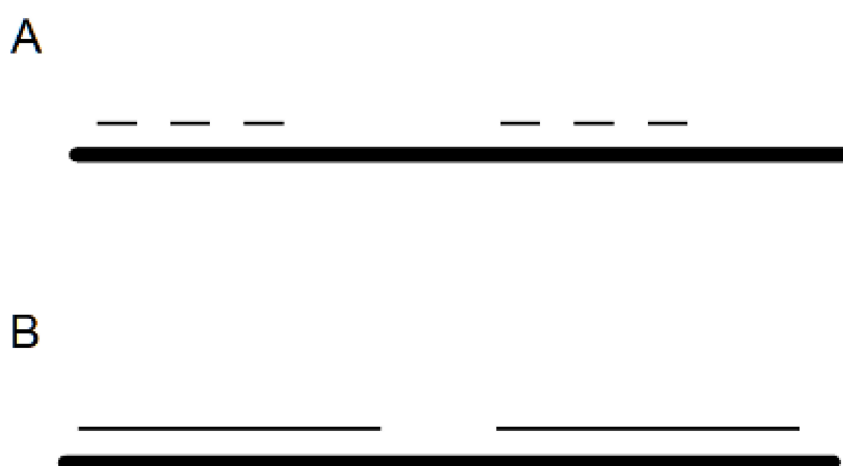


Figure 1. Tokenization of natural language samples. Each token may be assigned to a word or a string of words in a document. Generation of a text sequence by deep learning is dependent on the tokenization procedure. (A) The contiguous line represents a sequence of words as they appear in a document. Above this line are dashed lines which are tokens that correspond to individual words in the document. (B) Same as (A), except instead of subwords, the longer dashed lines represent a larger sequence of text. Therefore, each dash is a token that corresponds to many words in a document.

However, it is not certain that complex logic is easily captured by the above method. It is possible that training on translations of natural language samples to a logical format would increase performance of the large language models. Another possibility is that the tokenization process should include the higher-order features of natural language that correspond to logic, and other features that have a property of “meaning” in the context of cognition.

4. Large Language Models and Society

Deep learning and large language models are areas of rapid technological advancement, leading to impacts on society at this time [51]. As with the other important technologies and industries, there is competition across all levels of human society, from the blocs of nations, to the research centers in corporations, to the technology workers [52]. This activity occurs in many areas, including in the development of deep learning architectures, the design of dedicated computer processing units, investment in commercialization, and methods of trade to encourage development while discouraging competitors. Reflections and opinions on these activities are more easily heard in the form of reports by news media and other kinds of public communication. These methods of communication may range from informational on scientific discoveries to strong persuasion for increasing public support for a governmental policy [53].

Deep learning models are already impacting scientific research and investigations into natural phenomena [54,55]. In addition, there are current efforts for commercializing the large language models as a generalized method of searching for knowledge and to intelligently process multimedia data [56]. These efforts are supported by financial markets, news media, and policies at the national level. Not all technologies rapidly spread across society, but the software and algorithmic elements of deep learning are not bounded by regulation or cost, so advancement has and is expected to continue at a high rate. The ability to confidently predict the impact of science and technology is not possible, otherwise, the financial market of corporate shares in this industry would be deterministic; therefore, public discourse is unable to find a definite consensus on how to approach these technologies and their impacts on industry and government.

This entry is on the topic of logical reasoning for an increase in the interpretability of deep learning models, and for validating the generative output, such as the truth or falsehoods in natural language samples. Even though the construction of logical statements is not a universal necessity in human society, the use of logical reasoning in large language models has the potential for constructing knowledge with a higher degree of reliability than that constructed by traditional practices, such as rhetoric as bounded by historical constraint [6]. It is an open question on whether these traditional practices, as practiced in a society, are any more verifiable by a logic-based model than that generated by a robust large language model.

The problem of large language models in society is that information and communication is highly regulated and safeguarded by stakeholders involved in public policy and technological development. This would invite conflict between official statements of government and these models. It is also arguable that much of the communication in society is not based on verifiable evidence as in the natural sciences, so a large language model trained on logical reasoning can potentially measure and validate official speech, instead of relying on arguments by use of the rhetorical arts. While rhetoric may rely on methods outside logic and scientific evidence for a measure of truth, a model trained on logical reasoning is expected to be resistant to the persuasive methods of speech and propaganda.

Lastly, large language models are not incorporated into the current regime of laws and their interpretations, so as a result, it is not yet certain how deep learning will impact the assignment and development of intellectual property, such as in art and in music, and whether speech as generated by these models is protected by the lofty ideals of freedom of expression. These problems have arisen since these models have recently been released

for public use. While policy makers are forming opinions and finding solutions to open questions on regulation of this area of technology, the advancements in deep learning, as it pertains to both the natural sciences and natural language generation, are ongoing. As a confounding factor to the public discussion, large language models are becoming available that are effective yet contain fewer parameters; therefore, these models are potentially applicable for use in desktop computer systems [57].

The tension between policies at all levels and the deep learning technologies is ongoing. It is difficult to construct trade barriers to computer processors where the software technology is accelerating the computational efficiency of deep learning. Likewise, it may be difficult to regulate machine-generated speech and other generative output without limiting free expression by individuals in society. Lastly, there are existential questions on the unexpected use of deep learning technologies by corporations and governments. The most parsimonious answer to these questions is to use history as a guide. History is replete with examples of technology, its use in a society, and its use against competitors. While individuals vary in their personality and character traits, societal level institutions are often confined by rules and goals, such as a corporation in pursuit of profit and labor, or the allocation of wealth for increasing the standing of a nation among others [52]. It is expected that deep learning technologies will lead to a race for development of industry and national wealth, a race with a prize that may greatly exceed that of the marvels from the miniaturization of electronics [58].

5. Conclusions

The corpus of knowledge offers a great number of natural language samples to train a large language model by deep learning [1]. These models are capable of natural language prediction by this methodology. This predictive capacity is derived by training the model on big data and capturing the higher-level representations in the neural network [28]. It is also dependent on context size for finding associations across samples in this data. The larger the context size, the more easily the model is expected to capture the higher-level representations, in particular regarding the associative text samples that are not neighbors in the corpus (Figure 2). It is possible that scaling the size of these deep learning systems will continue to lead to increased capability and reliability in the generation of output [30]. However, it is also possible that the higher-order representations are not easily captured by this framework. This occurrence would arise if the data sample size would have to tend toward infinity for the purpose of capturing the complexity of propositional logic and other higher-order features inherent in natural language. An approach which considers symbolic and non-symbolic logic is a path for further development of large language models [23–25], and for training them to approach and then exceed those of their biological counterparts.

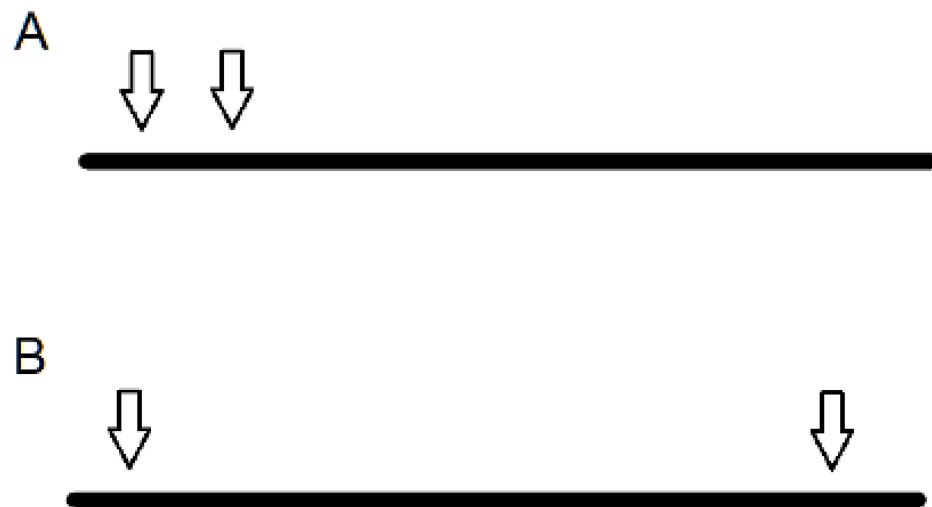


Figure 2. Associations between words in a document are frequently in close proximity, but they may also occur over a longer span in a document. (A) The contiguous line represents a sequence of words. Above this line are two arrows that point to sections of words in a document. Both these sections are associated and nearby one another. (B) Same as (A), except the associated sections are distantly located in the document.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Brants, T.; Popat, A.C.; Xu, P.; Och, F.J.; Dean, J. Large Language Models in Machine Translation. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), Prague, Czech Republic, 28–30 June 2007; pp. 858–867.
2. Hennig, W. Phylogenetic Systematics. *Annu. Rev. Entomol.* **1965**, *10*, 97–116. [CrossRef]
3. Scott-Phillips, T.C.; Kirby, S. Language evolution in the laboratory. *Trends Cogn. Sci.* **2010**, *14*, 411–417. [CrossRef] [PubMed]
4. Pinker, S.; Bloom, P. Natural language and natural selection. *Behav. Brain Sci.* **1990**, *13*, 707–727. [CrossRef]
5. Friedman, R. Tokenization in the Theory of Knowledge. *Encyclopedia* **2023**, *3*, 380–386. [CrossRef]
6. Waddell, W.W. *The Parmenides of Plato*; James Maclehose and Sons: Glasgow, UK, 1894.
7. Owen, G.E.L. Eleatic Questions. *Class. Q.* **1960**, *10*, 84–102. [CrossRef]
8. Merriam-Webster Dictionary. Available online: <https://www.merriam-webster.com/dictionary/rhetoric> (accessed on 6 April 2023).
9. The Britannica Dictionary. Available online: <https://www.britannica.com/dictionary/rhetoric> (accessed on 11 April 2023).
10. Rae, J.W.; Borgeaud, S.; Cai, T.; Millican, K.; Hoffmann, J.; Song, F.; Aslanides, J.; Henderson, S.; Ring, R.; Young, S.; et al. Scaling Language Models: Methods, Analysis & Insights from Training Gopher. *arXiv* **2021**, arXiv:2112.11446.
11. Traylor, A.; Feiman, R.; Pavlick, E. Can Neural Networks Learn Implicit Logic from Physical Reasoning? In Proceedings of the Eleventh International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023; (in review). Available online: <https://openreview.net/forum?id=HVoJCRLByVk> (accessed on 12 May 2023).
12. Evans, R.; Saxton, D.; Amos, D.; Kohli, P.; Grefenstette, E. Can Neural Networks Understand Logical Entailment? *arXiv* **2018**, arXiv:1802.08535.
13. Shi, S.; Chen, H.; Ma, W.; Mao, J.; Zhang, M.; Zhang, Y. Neural Logic Reasoning. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Online, 19–23 October 2020; pp. 1365–1374.
14. Horn, L.R.; Wansing, H. Negation. In *The Stanford Encyclopedia of Philosophy*; Stanford University: Stanford, CA, USA, 2015. Available online: <https://plato.stanford.edu/entries/negation> (accessed on 11 May 2023).
15. Aloni, M. Disjunction. In *The Stanford Encyclopedia of Philosophy*; Stanford University: Stanford, CA, USA, 2016. Available online: <https://plato.stanford.edu/entries/disjunction> (accessed on 11 May 2023).

16. Boole, G. *The Mathematical Analysis of Logic, Being an Essay towards a Calculus of Deductive Reasoning*; Macmillan, Barclay, & Macmillan: London, UK, 1847.
17. Leibniz, G.W. De Progressione Dyadica Pars I. 1679. In *Herrn von Leibniz' Rechnung mit Null und Einz*; Hochstetter, E., Greve, H.-J., Eds.; Siemens Aktiengesellschaft: Berlin, Germany, 1966.
18. Klement, K.C. Propositional Logic. Internet Encyclopedia of Philosophy. Available online: <https://iep.utm.edu/propositional-logic-sentential-logic> (accessed on 12 April 2023).
19. Russell, S. Unifying Logic and Probability. *Commun. ACM* **2015**, *58*, 88–97. [CrossRef]
20. Braine, M.D.; Reiser, B.J.; Rumin, B. Some Empirical Justification for a Theory of Natural Propositional Logic. *Psychol. Learn. Motiv.* **1984**, *18*, 313–371.
21. Garcez, A.D.A.; Gori, M.; Lamb, L.C.; Serafini, L.; Spranger, M.; Tran, S.N. Neural-Symbolic Computing: An Effective Methodology for Principled Integration of Machine Learning and Reasoning. *arXiv* **2019**, arXiv:1905.06088.
22. Yang, Y.; Zhuang, Y.; Pan, Y. Multiple knowledge representation for big data artificial intelligence: Framework, applications, and case studies. *Front. Inf. Technol. Electron. Eng.* **2021**, *22*, 1551–1558. [CrossRef]
23. Liang, P.; Potts, C. Bringing machine learning and compositional semantics together. *Annu. Rev. Linguist.* **2015**, *1*, 355–376. [CrossRef]
24. Hitzler, P.; Eberhart, A.; Ebrahimi, M.; Sarker, M.K.; Zhou, L. Neuro-symbolic approaches in artificial intelligence. *Natl. Sci. Rev.* **2022**, *9*, nwac035. [CrossRef]
25. De Raedt, L.; Dumancic, S.; Manhaeve, R.; Marra, G. From Statistical Relational to Neuro-Symbolic Artificial Intelligence. *arXiv* **2020**, arXiv:2003.08316.
26. Kant, I. *Critique of Pure Reason*; Weigelt, M., Translator; Penguin Classics: London, UK, 2003.
27. Friedman, R. A Perspective on Information Optimality in a Neural Circuit and Other Biological Systems. *Signals* **2022**, *3*, 410–427. [CrossRef]
28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*, 1–11.
29. Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Lee, Y.T.; Li, Y.; Lundberg, S.; et al. Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv* **2023**, arXiv:2303.12712.
30. Wei, J.; Tay, Y.; Bommasani, R.; Raffel, C.; Zoph, B.; Borgeaud, S.; Yogatama, D.; Bosma, M.; Zhou, D.; Metzler, D.; et al. Emergent Abilities of Large Language Models. *arXiv* **2022**, arXiv:2206.07682.
31. Schick, T.; Dwivedi-Yu, J.; Dessì, R.; Raileanu, R.; Lomeli, M.; Zettlemoyer, L.; Cancedda, N.; Scialom, T. Toolformer: Language models can teach themselves to use tools. *arXiv* **2023**, arXiv:2302.04761.
32. Efstathiou, V.; Hunter, A. Algorithms for generating arguments and counterarguments in propositional logic. *Int. J. Approx. Reason.* **2011**, *52*, 672–704. [CrossRef]
33. Lukins, S.; Levicki, A.; Burg, J. A Tutorial Program for Propositional Logic with Human/Computer Interactive Learning. *ACM SIGCSE Bull.* **2002**, *34*, 381–385. [CrossRef]
34. Ni, J.; Young, T.; Pandelea, V.; Xue, F.; Cambria, E. Recent Advances in Deep Learning Based Dialogue Systems: A Systematic Survey. *Artif. Intell. Rev.* **2022**, *56*, 3055–3155. [CrossRef]
35. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Zidek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [CrossRef]
36. Lin, Z.; Akin, H.; Rao, R.; Hie, B.; Zhu, Z.; Lu, W.; Smetanin, N.; Verkuil, R.; Kabeli, O.; Shmueli, Y.; et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science* **2023**, *379*, 1123–1130. [CrossRef]
37. Creswell, A.; Shanahan, M.; Higgins, I. Selection-Inference: Exploiting Large Language Models for Interpretable Logical Reasoning. *arXiv* **2022**, arXiv:2205.09712.
38. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language Models are Few-Shot Learners. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 1877–1901.
39. Chan, S.; Santoro, A.; Lampinen, A.; Wang, J.; Singh, A.; Richemond, P.; McClelland, J.; Hill, F. Data Distributional Properties Drive Emergent In-Context Learning in Transformers. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 18878–18891.
40. Beurer-Kellner, L.; Fischer, M.; Vechev, M. Prompting Is Programming: A Query Language for Large Language Models. *arXiv* **2022**, arXiv:2212.06094.
41. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Chi, E.; Le, Q.; Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *arXiv* **2022**, arXiv:2201.11903.
42. Taylor, R.; Kardas, M.; Cucurull, G.; Scialom, T.; Hartshorn, A.; Saravia, E.; Poulton, A.; Kerkez, V.; Stojnic, R. Galactica: A Large Language Model for Science. *arXiv* **2022**, arXiv:2211.09085.
43. Friedman, R. Themes of advanced information processing in the primate brain. *AIMS Neurosci.* **2020**, *7*, 373. [CrossRef]
44. Saharia, C.; Chan, W.; Saxena, S.; Li, L.; Whang, J.; Denton, E.L.; Ghasemipour, K.; Gontijo Lopes, R.; Karagol Ayan, B.; Salimans, T.; et al. Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding. *Adv. Neural Inf. Process. Syst.* **2022**, *35*, 36479–36494.
45. Floyd, J. Wittgenstein on Philosophy of Logic and Mathematics. *Grad. Fac. Philos. J.* **2004**, *25*, 227–287.
46. Hinton, G.E. Connectionist learning procedures. *Artif. Intell.* **1989**, *40*, 185–234. [CrossRef]
47. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [CrossRef]

48. Srivastava, A.; Rastogi, A.; Rao, A.; Shoeb, A.A.M.; Abid, A.; Fisch, A.; Brown, A.R.; Santoro, A.; Gupta, A.; Garriga-Alonso, A.; et al. Beyond the Imitation Game: Quantifying and extrapolating the capabilities of language models. *arXiv* **2022**, arXiv:2206.04615.
49. Fusi, S.; Miller, E.K.; Rigotti, M. Why neurons mix: High dimensionality for higher cognition. *Curr. Opin. Neurobiol.* **2016**, *37*, 66–74. [[CrossRef](#)]
50. Demortier, G. Revisiting the construction of the Egyptian pyramids. *Europhys. News* **2009**, *40*, 27–31. [[CrossRef](#)]
51. Tamkin, A.; Brundage, M.; Clark, J.; Ganguli, D. Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models. *arXiv* **2021**, arXiv:2102.02503.
52. Porter, M.E. The Competitive Advantage of Nations. *Harv. Bus. Rev.* **1990**, *68*, 73–93.
53. Lippmann, W. *Public Opinion*; Harcourt, Brace and Company: New York, NY, USA, 1922.
54. Choudhary, K.; DeCost, B.; Chen, C.; Jain, A.; Tavazza, F.; Cohn, R.; Park, C.W.; Choudhary, A.; Agrawal, A.; Billinge, S.J.; et al. Recent advances and applications of deep learning methods in materials science. *NPJ Comput. Mater.* **2022**, *8*, 59. [[CrossRef](#)]
55. Meher, S.K.; Panda, G. Deep learning in astronomy: A tutorial perspective. *Eur. Phys. J. Spec. Top.* **2021**, *230*, 2285–2317.
56. Liu, Y.; Han, T.; Ma, S.; Zhang, J.; Yang, Y.; Tian, J.; He, H.; Li, A.; He, M.; Liu, Z.; et al. Summary of ChatGPT/GPT-4 Research and Perspective Towards the Future of Large Language Models. *arXiv* **2023**, arXiv:2304.01852.
57. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.A.; Lacroix, T.; Roziere, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. LLaMA: Open and Efficient Foundation Language Models. *arXiv* **2023**, arXiv:2302.13971.
58. Newton, A.; Dhole, K. Is AI Art Another Industrial Revolution in the Making? *arXiv* **2023**, arXiv:2301.05133.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.